UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

May 3, 2021

DSSD 2020 DECENNIAL CENSUS MEMORANDUM SERIES #J-05

MEMORANDUM FOR Albert E. Fontenot, Jr.
Associate Director, Decennial Census Programs

From: Patrick J. Cantwell
Chief, Decennial Statistical Studies Division

Prepared By: Julianne Zamora
Chief, Estimation Branch
Decennial Statistical Studies Division

Subject: 2020 Census Count Imputation - High-Level Overview

This memorandum provides a high-level overview of the count imputation process for addresses in the housing unit universe of the 2020 Census. It does not cover count imputation for group quarters.

For further information, contact Julianne Zamora 301-763-7994,
Andrew Keller at 301-763-9308, or Timothy Kennel at 301-763-6795.

cc:

Deborah M. Stempowski
Assistant Director, Decennial Census Programs, Operations and Schedule Management
Acting Chief, Decennial Census Management Division

Karen Battle
Division Chief, Population Division

John M. Abowd
Associate Director for Research and Methodology and Chief Scientist

[This page is intentionally left blank.]

1 Count Imputation

In this memorandum, we provide a high-level overview of the 2020 Census count imputation process for housing unit addresses. Count imputation for group quarters was a separate activity and not included in this document. Count imputation occurred after the 2020 Census field data collection to ensure that every housing unit address had a final status and population count (if applicable). The enumeration portion of the 2020 Census occurred in two primary stages:

1) Self-response
2) Nonresponse Followup (NRFU)

Census enumeration attempts to obtain complete counts of the population and housing units in the United States and Puerto Rico. To do this, each address must be assigned a final housing unit status of **occupied**, **vacant**, or **delete** (i.e., nonexistent). If the status is occupied, then the address must also have a population count greater than zero.

At the end of the 2020 Census enumeration operations, some addresses were unresolved, meaning we lacked sufficient response information to assign them a final housing unit status or population count. Thus, count imputation had two functions:

1) Fill in missing housing unit status.
2) Fill in missing population counts for housing units that are a) known to be occupied but household size is unknown, or b) imputed to be occupied.

An address that required count imputation fell into one of three **imputation types**:

1) **Status Imputation** - The address may or may not represent a valid housing unit. It could have been an occupied unit, a vacant unit, or a delete. Some of these addresses were not found during census operations. Some addresses were assigned to status imputation due to conflicting or insufficient information from NRFU or other census operations. Finally, some addresses were assigned to status imputation if all people at the address were already counted elsewhere.
2) **Occupancy Imputation** - The housing unit was known to exist, but could have been occupied or vacant.
3) **Household Size Imputation** - The housing unit was known to be occupied, but the population count was unknown.

For the 2020 Census, count imputation used the nearest-neighbor hot deck method. In the nearest-neighbor hot deck, the unresolved housing unit status or population count is imputed using the values from a resolved “donor” record. The donor record is the nearest resolved address using an address sort that shares specified characteristics with the unresolved record.

2 Cell Assignment

The first part of the process was for the Census Bureau to assign all addresses (both resolved and unresolved) to groups referred to as *imputation cells*. Addresses in each cell share characteristics that are related to the housing unit status and population count. These characteristics were derived from the Master Address File (MAF), NRFU operation, and administrative records (AR) (if they exist). In each state, self-response addresses were separated from other types of responses (such as enumerator responses from NRFU). Self-response addresses were split into nine cells within each state based on the AR count. For other types of responses, the seven variables in Table 1 were combined to create 86 imputation cells in each state. The full cross-classification of these seven variables was not used. The cells were ordered so that nearby cells share similar distributions of household status and population count based on responses from the 2010 Census.

Table 1: 2020 Count Imputation Variables and Descriptions

Variable	Description
Nearest-Neighbor Household Type	Classified each address based on the housing unit status of the address's nearest resolved neighbor (e.g., occupied, delete, vacant).
MAF Unit Status	Classified each address as valid living quarters or not (e.g., demolished, delete, duplicate).
MAF X-Type Flag	Classified the address as likely delete, likely vacant, or other.
Spring Delivery Sequence File (DSF) Flag	Classified the address as residential or other (e.g., commercial or not on the DSF).
NRFU Proxy Type	Classified the address as having an unknown proxy respondent (no name, etc.) or other respondent (e.g., proxy with known respondent, no proxy).
Undeliverable as Addressed (UAA) Reason Code	Classified the address into one of three categories: No such number, all other UAA codes, or no UAA code (includes addresses with no mailing).
Administrative Records (AR) Count	Classified the address based on the address's population count according to administrative records (not found in AR or 0 AR people, 1 – 9+).

3 Nearest-Neighbor Hot Deck

The addresses in each state were sorted first by the imputation cell identifier, then by geography. For geography, the sort included area census office (ACO), county, tract, basic collection unit (roughly the size of a block) and, finally, the walking sequence within the basic collection unit.

For each unresolved address, the nearest resolved address preceding it on the sorted file was generally the donor. If there wasn't a donor preceding the unresolved address in the same tract and cell, the search continued at the bottom of the list within the same tract and cell. If there wasn't a donor within the tract, the search extended to the addresses in the same ACO in the cell in a similar manner, starting with the preceding tract in the same ACO. If there wasn't a donor within the same ACO, the search extended to the addresses within the same state and cell. Finally, if there wasn't a donor within the same cell in the state, the search for the donor

began with the last address in the list for the preceding cell, within the same state. There was no limit on the number of times the same donor could be used.

The same cell definitions and sort were used for status imputation, occupancy imputation, and household size imputation. The resolved delete addresses were excluded from occupancy and household size imputation. Resolved vacant addresses were excluded from household size imputation.

The sorting method ensured that the unresolved address was geographically close to the donor and shared similar characteristics with it. The quality of the imputations is directly related to the strength of the spatial correlation between unresolved addresses and resolved unoccupied addresses. That is, the nearest-neighbor hot deck works better if unresolved addresses are more likely to be adjacent to an address that is resolved as vacant or delete. Further, for occupied addresses, the quality of the imputations is related to the spatial correlation in the population count for addresses that are resolved as occupied.

Any questions should be directed to Julianne Zamora, Andrew Keller, or Timothy Kennel (Assistant Division Chief, Statistical Methods, Decennial Statistical Studies Division) at 301-763-6795.

4 References

Rowe, C., Keller, A., Heim, K. (2020). “Decennial Response Processing System (DRPS) Technical Specifications Document: Census Unedited File (CUF) Count Imputation Computer Software Requirements,” DSSD 2020 Decennial Memorandum Series J-03.